HELMET-MOUNTED DISPLAY COMPUTER MODEL
AND RESEARCH VISUALIZATION TOOL

T.H. Harding*, J.S. Martin and C.E. Rash
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

ABSTRACT

As the Army increases its reliance upon helmet mounted displays (HMDs) or head-up displays (HUDs), it is paramount that displays are developed that meet the operational needs of the warfighter. During the development cycle, questions always arise concerning the operational requirements of the HMD. These include questions concerning luminance, contrast, color and resolution. To provide intelligent answers to these operational questions, a method has been devised to evaluate these issues. Integral to this method is a HMD simulation model. The model allows for contrast correct visualizations of see-through imagery. The imagery consists of symbology/situational maps overlaid over natural backgrounds. Using visual psychophysical procedures, observers judge the quality of the symbology for a range of luminance and background conditions. The simulated images were analyzed and statistical correlates were developed that could relate to the observer's ratings. Metrics were developed that could help predict the operational luminance requirements for HMD or HUD symbology.

1. INTRODUCTION

For symbology to be viewed in a see-through helmet mounted display (HMD) or head-up displays (HUD), the luminance of the symbology must be sufficient to distinguish it from the see-through background. When the contrast of the transparent symbology is sufficiently high, the symbology appears as an overlay on the ambient scene. The luminance requirements in order for an HMD or HUD to be usable in an operational environment must take into consideration the type of displayed imagery (e.g., symbology, situational maps, target sights), the tasks (e.g., targeting, navigation, obstacle avoidance), the operational setting (e.g., day/night, terrain features), additional hardware (e.g., visors, windscreens, laser protection), and other considerations.

As the Army increases its reliance upon augmented vision devices, such as HMDs and HUDs, it is paramount that the devices developed meet the operational needs of the warfighter. Historically, program managers and combat developers have estimated performance specifications based upon simplifying approximations regarding optical performance. These specifications may lack operational validity. To provide more accurate and meaningful performance specifications, we developed a computer model that simulates the performance of HMDs/HUDs and provides a means of quantifying performance requirements (Harding et al., 2002).

2. HMD MODEL

The model consists of a graphical user interface (implemented in Microsoft Excel 2003) using Microsoft's Visual Basic for Applications. Excel offers a tremendous advantage when working with arrays and implementing formulas quickly. Its database capabilities provide easy access to data, while its charting and graphing support make most tasks required of the model routine. Once model parameters are selected and the formulas processed, the resulting data are transferred to an executable program that processes images using the data calculated from the spreadsheet.

Incorporated into the model is a database of the transmission spectra of aircraft windscreens, visors, protective masks, laser protection devices, personal eyewear, and a few HMDs. Likewise, the database includes the emission spectra of several cockpit displays, lasers, and HMDs. The model easily allows for the inclusion of custom or user defined hardware spectra.

Figures 1 and 2 provide a graphic of the basic model. In Figure 1, outside imagery is filtered first by an aircraft windscreen. The windscreen-filtered image can be filtered in turn by a visor, an HMD combiner lens, a protective mask, and eyewear. This progression of filters results in an image of the outside world that can be greatly altered in color appearance and in luminance. Figure 2 depicts the filtering process of cockpit imagery, such as the emission spectra of multifunction displays, reflective spectra of cockpit instruments, etc. The transmittance spectra of windscreens, visors, and protective masks, as well as items of personal eyewear, are fairly flat except for items where laser protective coatings or dyes have been used. As a general rule, the major alterations to the color of outside and inside imagery occurs as a result of filtering by the HMD, where the HMD's combiner lens has coatings to maximally reflect the HMD source emission spectra.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Helmet-Mounted Display Computer Model And Research Visualization Tool	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Rucker, AL 36362	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**8**	

The ability to simulate imagery, as viewed through the various filters, requires a color space transformation from spectra to RGB. The spreadsheet transforms the spectra into CIE XYZ and then from XYZ to RGB. For the case of transmission spectra, RGB scalars or coefficients are obtained that allow filtering of RGB imagery. HMD and display emission spectra are transformed into an RGB color. As computer models cannot handle the range of luminance encountered in the real world, the model produces Gamma-corrected luminance contrast for the monitor being used. This approach is quite acceptable; since, over most of the photopic range (i.e., above about 10fL), there are only small differences in visual contrast sensitivity (Barten, 1990, 1992).
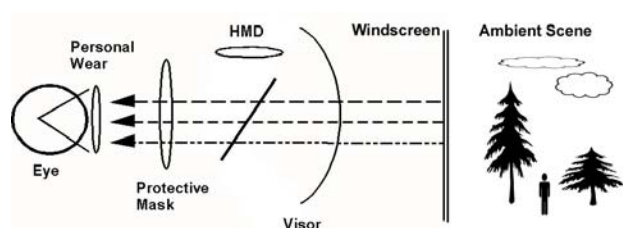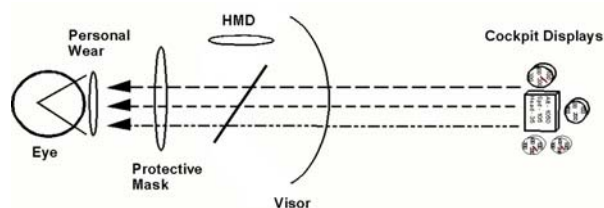

Figure 1. Basic model for ambient imagery.

Figure 2. Basic model for cockpit displays.

2. METHOD

Relating visual performance to the output of the model requires the production of hundreds of computer generated images. The imagery generated shows symbology overlaid over natural images. The contrast of the symbology to the background is a function of the parameters chosen for the symbology. In a controlled experiment, observers evaluated the quality of the symbology in each of the images.

2.1 Simulations

Ten background images were used to represent a range of background scenes and lighting conditions (Figure 3). Eight of the images were natural scenes representing landscapes, seascapes, ground clutter, and sky. The two other images were a uniform field (representing the simplest background) and a grayscale image composed of abrupt changes in contrast (representing a difficult background). Abrupt changes in contrast are characteristic of high spatial frequency content.

During the computer simulations, the symbology was set to a luminance of 1000 foot-Lamberts (fL), with the red, green, and blue lasers matched, such that at equal gray levels (levels of 0 to 255) the color was a shade of gray. The simulations used the transmission spectra of an UH-60 Black Hawk windscreen and a Gentex tinted visor (Figure 4). We constructed the transmission spectra for a triple-notched HMD combiner lens, with the notches corresponding to the three laser wavelengths (Figure 4). For each background image, simulations were conducted for a peak ambient luminance of 500 to 10,000fL in 500fL increments. These simulations were repeated with the tinted visor removed. A total of 400 simulations were conducted. Of these, a subset of 200 was chosen for this study. As the tinted visor is essentially a neutral density filter, the simulations chosen for inclusion were those that covered the entire range of contrast expectations. Obviously, it is not possible to make an actual assessment of symbology, using a CRT monitor, over the range of luminances used in these simulations. Because of this, these simulations produce images in which all have about the same average luminance but differ in their contrast. As an example, when a 10,000-fL simulation is performed, the relative luminances are calculated and then reproduced at the ambient luminance of the computer monitor (about 14fL). The contrast between symbology and background is the same as it would be at 10,000fL, only it is displayed at the average monitor luminance.

The average at-the-eye luminance for ten background images are presented in Table 1 for the 10,000fL conditions. Given the overall hardware transmission spectra and the differences in luminance for the ten background images, average luminance at the eye would range from a low of 7fL for the Ground Clutter 2 image (500fL peak with the tinted visor) to a high of 6,198fL for the uniform background (10,000fL peak with no visor). All model calculations were conducted using a monitor gamma of 2.2. The luminance response curve of the Sony Trinitron monitor used in this study could be well described by a gamma of 2.2.

2.2 Symbology

A symbology image was developed in-house where the smallest letters subtended 37.5 arc minutes. The gaps between the small letters were 7.5 arc minutes; this equates to a Snellen letter size of 150. This exceeds the Army's recommended symbology Snellen size of about 100 (MIL-STD-1787). The larger letter size was chosen to minimize resolution as a factor in an observer's ratings. The symbology image can be seen in Figure 5.
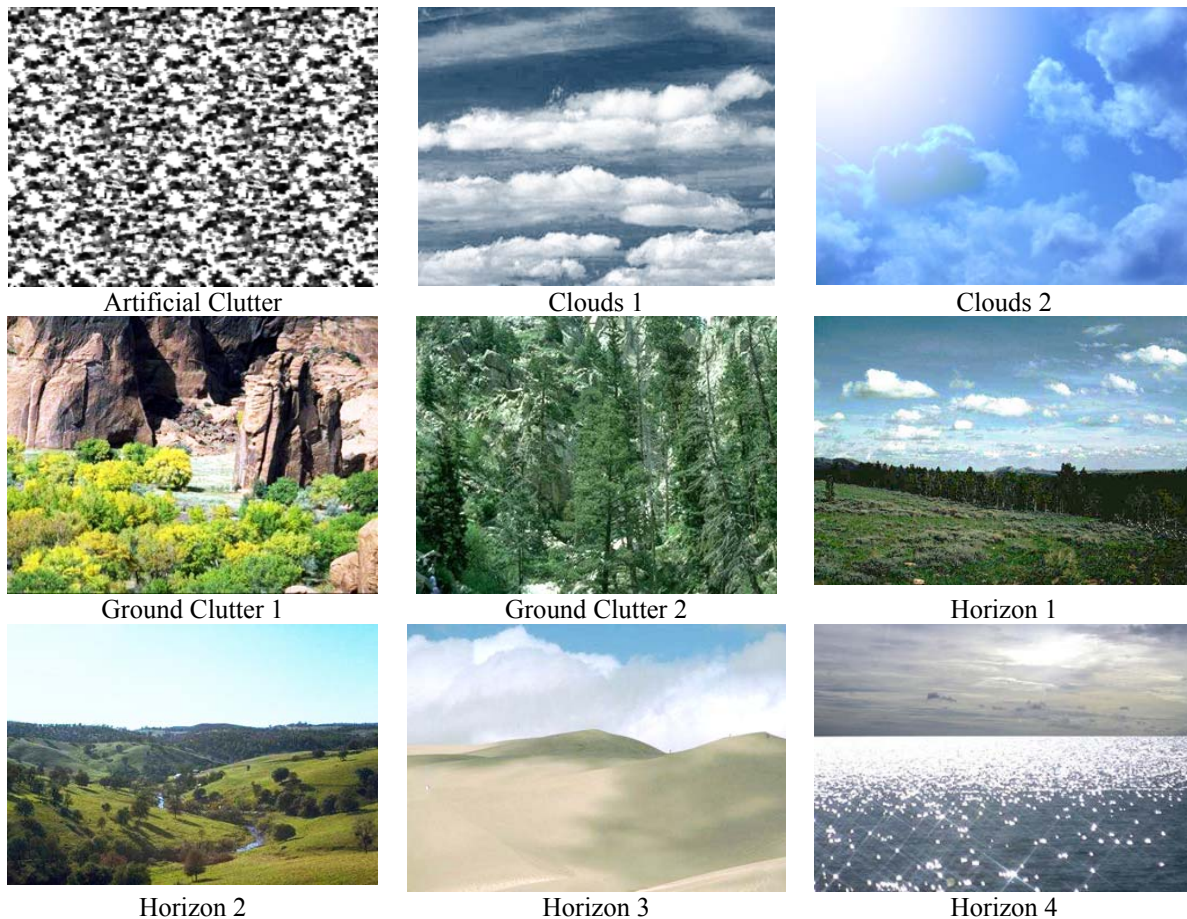
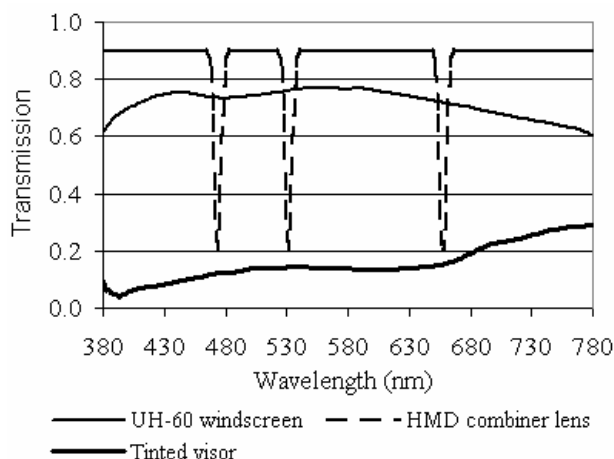Figure 3. Nine of the ten background images used in the study. The tenth image was a uniform grayscale image.

Figure 4. Transmission spectra for the hardware used in the modeling.

2.3 Evaluation Scale

To evaluate the quality of the symbology, a scale of 1 to 7 was developed. The observer assigned a rating of 1 to 7 for each of the 200 simulated images. To assist observers, guidelines were developed that differentiate between the seven ratings. The rating scale developed is shown in Table 2. These instructions, along with symbology examples, were given to the observer prior to testing. In addition, the rating instructions were available anytime during testing via a click of a mouse. In our scale, a rating of 4 would be considered the lowest quality symbology that could be used operationally. In the description of the rating scale, the term contrast was used to describe perceived contrast and not physical contrast as these were instructions to observers. Perceived contrast is affected by luminance contrast and background texture (Barten, 1990). Increases in background clutter reduce the perceived contrast.

2.4 Display

The Sony Trinitron 17-inch CRT monitor had a native screen resolution of 1280 by 1024 pixels. With a horizontal field-of-view of 40 degrees, the symbology characters yielded the required angular subtense of 37.5 arc minutes at a viewing distance of 22 inches. To assure that observations were made at this viewing distance, a chin rest was used and adjusted for each subject's eye

position and distance to the screen. To obtain the correct angle, the 640 by 480 images were doubled yielding images of 1280 by 960 pixels. The monitor's additional 64 pixels at the bottom of the screen allowed space for display of the rating scale and other buttons.

Table 1. Average at-the-eye luminances for the 10,000 fL peak-luminance conditions.

Background Image	Scale Factor	Tinted Visor	No Visor
Artificial clutter	43.33%	411	2686
Clouds 1	42.93%	407	2661
Clouds 2	60.78%	576	3767
Ground clutter 1	31.03%	294	1923
Ground clutter 2	14.35%	136	889
Horizon 1	31.30%	297	1940
Horizon 2	42.19%	400	2615
Horizon 3	61.24%	580	3796
Horizon 4	44.77%	424	2775
Uniform	100%	948	6198

Figure 5. This symbology image was overlaid over the background.

Table 2. Rating scale and descriptions.

Rating	Quality	Description of rating
7	Excellent	All letters and symbols easily seen at high contrast
6	Very Good	All letters and symbols easily seen with good contrast.
5	Good	All letters and symbols are easily seen with reduced contrast.
4	Adequate	All letters and symbols can be deciphered with a little difficulty
3	Poor	Letters and symbols can barely be detected and some letters or symbols are very difficult to see.
2	Not adequate	Some of the letters and symbols cannot be seen.
1	Not usable	Difficult to recognize that symbology is present.

2.5 Observers and Sessions

Twenty volunteers from the USAARL workforce served as observers. All observers were required to pass an intermediate-field acuity test (distance of 22 inches) and a Farnsworth 15-Hue color vision test. There were no gender or age specific requirements. A session consisted of evaluating all 200 images (which were presented in random order). Breaks were programmed at the end of 50, 100, and 150 image presentations. On average, each trial took about 30 minutes to complete. Each subject completed four sessions, with only the last three sessions used for data analysis, as the first session was used for observers to develop judgment criteria.

2.6 Observer Reliability

To evaluate an observer's reliability, ratings on each image were compared for the three sessions. If the ratings for an image were the same, one point was awarded. If any two of the three ratings differed by more than 1 (a judgment error), six points were deducted. For each of the 10 backgrounds, the ratings for the 20 simulated images were likewise compared. If a lower luminance image had a lower rating than the next higher luminance image, then a point was deducted for each occurrence (an order error). Using this scheme, each observer received a reliability score. The scores ranged from a high of 96 to a low of -328. A grade of A through F was assigned to the distribution of scores. Table 3 contains the distribution and the grade scores. A grade of C was deemed sufficient for that observer's data to be included in the statistical analysis. Hence, the data from only 7 of the 20 observers were used in this study.

Table 3. Distribution of observer reliability scores.

Grade	Rating Range	Occurrences
A	90 and above	3
B	80 to 89	1
C	60 to 79	3
D	0 to 59	4
F	less than 0	9

2.7 Image Analysis

To make analysis of simulated images easier, a symbology image containing 20 square patches (each subtending about 1.6°) was processed in exactly the same fashion as the symbology image used for observer ratings. In Figure 6a, symbology is overlaid over the Horizon 1 background image. In Figure 6b, the block symbology image is overlaid on the same Horizon 1 background image. Both images were identically processed and have the same contrast. Statistical comparisons were made for each of the 20 square areas and the symbology-free rectangular areas of the same size and adjacent to them. In Figure 6, two examples are shown. The symbology patch B is compared to the non-symbology patch A, and

the symbology patch C would be compared to the non-symbology patch D. In this way, local contrast and signal-to-noise measures and other statistics were made for each of the 200 images. Image statistics then were compared to observer ratings.

3. RESULTS

By comparing observer ratings and certain statistical properties of the images, important correlates may be found that can provide information about the visual requirements for symbology. Multiple measures of contrast were determined for each of the 200 images. For example, the average Michaelson contrast was calculated. All measures of contrast were averages, where the average luminances of each of the 20 patches were compared with luminances from their adjacent patches (Figure 5). Separate average luminances (L) were obtained from the R, G, and B components of the image and these were combined for an overall average where
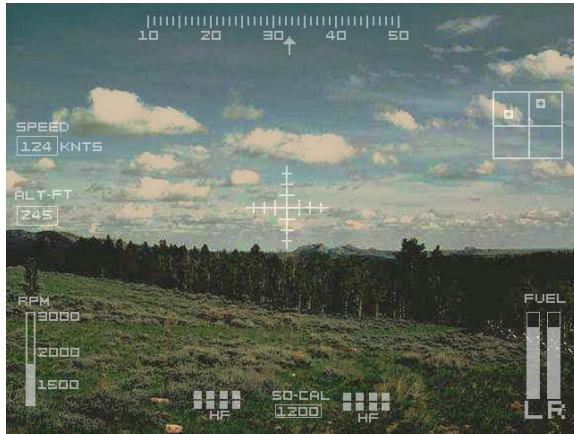
$$L = [(3_{x,y} \ (R_{x,y} + G_{x,y} + B_{x,y})]) \ n$$

and R, G, and B are luminance components calculated from the rgb grayshade values (0 to 255) where $R = (r/255)^{2.2}$, $G = (g/255)^{2.2}$, $B = (b/255)^{2.2}$, and n is the total number of pixels in each summation. To calculate Michaelson contrast $[(L_{max}-L_{min}))(L_{max}+L_{min})]$, L_{max} is L of the symbology patches and L_{min} is L of the adjacent patches.

In Figure 7, for each of the 200 images, Michaelson contrasts are plotted against ambient luminance. As expected, the contrasts from all ten background images converge to a single curve. The slight perturbations are due to the method of calculation. The contrasts are calculated from the 20 pairs of patches, and the average luminance is calculated from the entire image. The Uniform Field has the lowest contrast values, whereas the Ground Clutter 2 image has the highest. In Figure 8, Michaelson contrasts are plotted against the average observer ratings for each of the 200 images. Note the data diverge for contrasts below about 80%, indicating Michaelson contrast is a poor indicator of symbology

Figure 6a. Horizon 1 image with overlaid symbology.

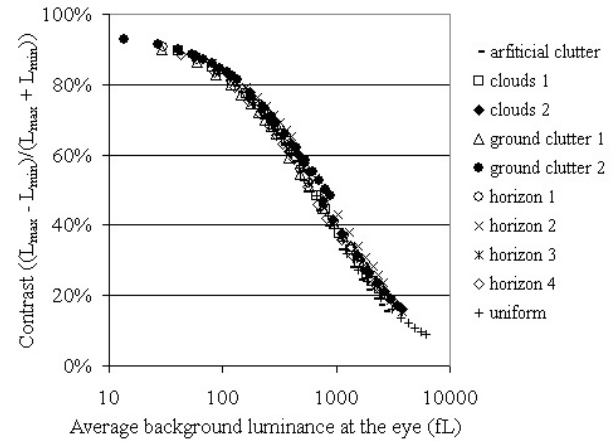Figure 6b. Horizon 1 image with overlaid block symbology.

Figure 7. Michaelson contrast as a function of simulated background luminance for the ten background images.
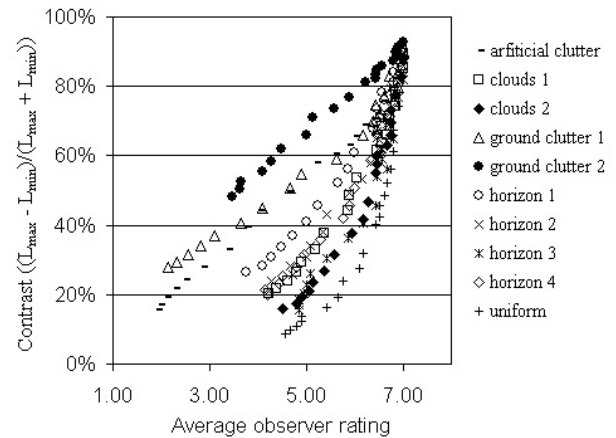

Figure 8. Relationship between Michaelson contrast and average observer ratings for the ten background images.

ratings. It also is interesting to note that ratings were higher for the Uniform Field, even though the Uniform Field has the lowest contrasts. This suggests that the complexity of the background has a greater influence on observer ratings than does average contrast.

3.1 Uniform Backgrounds

Developing predictors for the quality of symbology as a function of natural backgrounds first requires description of the quality of symbology against a uniform background. Since the symbology patches and the background patches are geometrically separated, it is easy to quantify the average luminance of the symbology (L_{sym}) as $L_{max} - L_{min}$. In Figure 9, the average observer ratings for the uniform field are plotted as a function of L_{sym}) (0.1 * L_{min}). The three highest ratings are ignored since the values were at or near the maximum observer rating of 7. The constant 0.1 is used so the ordinate would equal 1.0 at an average observer rating of 4.0 which is the lowest quality rating that would be considered operationally viable (see ratings instructions in Table 2). The exponential curve fitted to the data in Figure 9 has a predicted observer rating of 4.07 when the ordinate equals 1.0. The rationale for developing equations that would equal 1.0 near an average observer rating of 4.0 is that it simplifies calculations of required symbology (see Discussion).
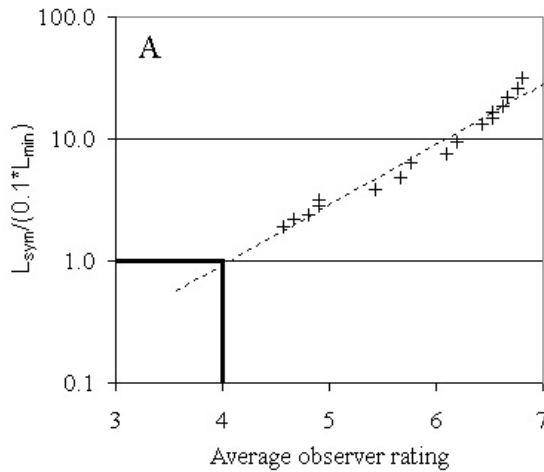

Figure 9. Plotting of the observer ratings for the uniform background images as a function of $L_{sym}/(01*L_{min})$. The three highest data points are not plotted. The dashed line is an exponential fit to the data ($y = 0.0097e^{1.14x}$) and has been extended to show its intersection at an average observer rating of 4.0 (bold lines).

Based upon the above data for uniform backgrounds, the minimum luminance of L_{sym} must satisfy the condition:

$$L_{sym} \ni 0.1 * L_{min.} \qquad (1)$$

As the observer ratings were higher than 4.0, the extrapolation to a rating of 4.0 is based entirely upon the exponential fit to the data. In the data presented below for complex backgrounds, an exponential curve fit to the data provided exceptional fits for all background conditions and therefore some confidence can be placed in the extrapolation.

3.2 Complex Backgrounds

To evaluate the complexity of the background images, the standard deviation of each of the background patches was calculated, and the average of these deviations is represented by SD_{min}. As with the luminance measurements, SD_{min} is the average of the average R, G, and B standard deviations calculated separately. In Figure 10, the average observer ratings for all ten background images are plotted against an ordinate of L_{sym}) [(0.1*L_{min}) + (1.42*SD_{min})]. The SD_{min} scalar of 1.42 is the average of the scalars fitted to the regressions in Figure 11 for the complex backgrounds (see discussion below). These data are more tightly grouped than the Michaelson contrast data shown in Figure 8, demonstrating that the variability of the background indeed influences the quality of symbology.
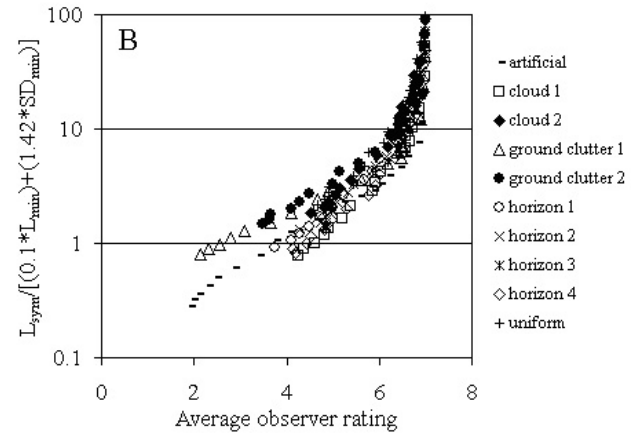

Figure 10. Plotting of average observer ratings for all backgrounds as a function of $L_{sym}/[(0.1*L_{min})+(1.42*SD_{min})]$.

To evaluate each background image separately, observer ratings (below 6.8) were fit with an exponential curve where the scalar for SD_{min} was adjusted so that at an observer rating of 4.0, each curve would approximate 1.0. Figure 11 shows exponential curves that were fit to each set of data. All of the exponentials had a product moment correlation value of 0.96 or higher.
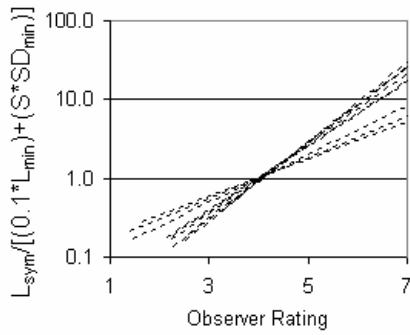
Figure 11. Exponential curves fit to the average observer ratings for each of the ten background images. S in the ordinate equation stands for the SD_{min} Scalar.

4. DISCUSSION

In this study, observers rated the quality of symbology, overlaid over natural background images, on a scale of 1 to 7. Observer ratings were compared with statistical measures derived from images of block symbology (Figure 6b) that were processed in identical fashion to the images used in the visual test. Luminance contrast (i.e., Michaelson contrast) was a poor predictor of observer ratings. For example, observer ratings were higher for the uniform background condition, even though the average contrast was significantly lower. Compare this to the Ground Clutter 2 image that had the highest average contrasts, but produced the third lowest observer ratings.

The present study revealed that the complexity or standard deviation of the background image was of paramount importance in determining the luminance requirements for symbology. Surprisingly, the luminance of the background image was of less importance. On average, the following equation summarizes the results:

$$L_{sym} \geq [0.1+(1.42*SD_B)] * L_B \qquad (2)$$

where L_{sym} is the luminance of the symbology, and SD_B is the standard deviation of the ambient background. For a uniform background, the minimum symbology luminance is only 10% of the background luminance. When the standard deviation of the background is 70%, the minimum symbology luminance increases to 110% of the background luminance (an eleven fold increase).

4.1 Background Complexity

In this study, standard deviations provided a rather good estimate of background complexity, but the metric is not perfect. For example, in Figure 12 two grayscale images are shown with each image having about the same standard deviation and about the same average contrast (text/background). Overlaid text (text added to background) can clearly be seen in the image on the left but is difficult to decipher any characters on the image on the right. The two images differ in their spatial frequency content. The gradient background is characterized by low spatial frequencies, whereas the random background is characterized by high spatial frequencies.
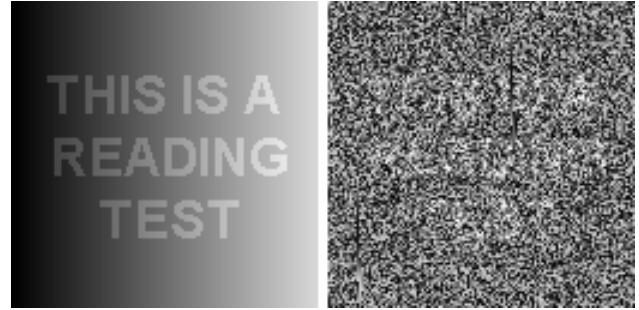

Figure 12. Transparent text overlaid over two images of near identical standard deviations. The image on the left is a gradient.

In this study, standard deviations were not calculated for the whole background image, but for small background patches in areas where symbology was overlaid (Figure 6b). The square patches subtended an angle of 1.56 degrees. As patch size is reduced, lower spatial frequency modulation is also reduced. It seems that a better metric would characterize background clutter only where the symbology is or could be overlaid. Hence, to have better predictive capability, measures must be developed that provide an elemental analysis of background areas that approximate the size of symbology characters.

4.2 Predictions

Equation 2 provides an average fit to the data collected here. However, when considering minimum luminance requirements for symbology, it is best to consider the worst case and not the average of the eight natural background images. The worst case condition had a standard deviation scalar of 2.95. Likewise, the highest standard deviation, for one of the eight natural scene images, was 74.9%. Using these two numbers yields a minimum luminance for symbology of $2.3*L_B$.

For purposes of this exercise, 5,000fL would be considered a high out-the-cockpit ambient luminance. Using the $2.3*L_B$ formula, and the transmission spectra from Figure 4 for the UH-60 windscreen, the Gentex tinted visor, and the modeled HMD, yields an HMD emissions requirement of 1,006. Without the tinted visor, a seven-fold increase in luminance requirement to 7,358fL is required.

To characterize the luminance requirements for situational maps and other more complex imagery is a much more difficult task than reading symbology. A symbology set is characterized by the quasi-stationary arrangement of symbols and indicators. For example, engine RPM would generally be located at the same location all of the time. It would not change from the left side of the display to the right side randomly. There is general positional certainty for almost all elements of the symbology set. Once the location and representative symbology set becomes committed to memory, there is very little uncertainty in the information conveyed. This cannot be said for situational type displays. These displays, by their very nature, change from day to day and even from minute to minute. This perceptual uncertainty translates to increased requirements for image contrast. That is, greater image quality is required for situational type displays than for symbology. Preliminary data from our laboratory confirms this increased luminance requirement for situational type displays.

CONCLUSION

By using an HMD model to process contrast correct visualizations of symbology overlaid over natural backgrounds, we have found that the complexity or standard deviation of the background image was of paramount importance in determining the luminance requirements for symbology. Surprisingly, the luminance of the background image was of less importance. From this data, we developed equations that can predict the quality of symbology against natural backgrounds. By extending this analysis, we developed a tentative luminance requirement for HMDs based on simulated operational conditions.

This study should be considered a preliminary study of luminance requirements for symbology, as we have only tested static, white symbology against static background images. With the background in motion, the symbology should stand out more and thus the minimum luminance requirement presented here are likely to be too stringent. In addition, variables such as vehicle vibration, color contrast, and soldier stress will complicate the calculations presented here. Likewise, complex overlaid imagery such as situational maps, will increase observer uncertainty and thus the HMD luminance requirements would be higher. Our research plan is to examine these and other issues.

REFERENCES

Barten, P.G.J., 1990: Evaluation of subjective image quality with the square-root integral method. *J. Opt. Soc. Am.*, 7, 2024-2031.

Barten, P.G.J. 1992: Physical model for the contrast sensitivity of the human eye. *Human Vision, Visual Processing, and Digital Display III, Proc. SPIE*, 1666, 57-72.

Harding, T.H., Klymenko, V., Martin, J.S., and Rash, C.E., 2002: An operational helmet-mounted display model: prediction of visible grayshades and see-through spectral data, *SPIE Helmet- and Head-Mounted Displays VII*, 4711, 290-298.

MIL-STD-1787, (2002). Appendix E, "Handbook for rotary wing aircraft display symbology", August 31.